

# Komparasi *K-Means Clustering* Dan *Complete Linkage* Dalam Pengelompokan Penyaluran Pinjaman *Financial Technology*

Akmal Fikri<sup>1\*</sup>, Benedika Ferdian Hutabarat<sup>2</sup>, Ulfa Khaira<sup>3</sup>

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi  
Jl. Jambi - Muara Bulian No.KM. 15, Jambi, Indonesia  
akmalfikri908@gmail.com<sup>1</sup>, benedika@unja.ac.id<sup>2</sup>, ulfa.ilkom@gmail.com<sup>3</sup>

Submitted : 21/07/2023; Reviewed : 03/09/2023; Accepted : 17/09/2023; Published : 31/10/2023

## Abstract

This study aims to compare the effectiveness and efficiency of the K-Means Clustering and Complete Linkage methods in classifying loan distribution by Financial Technology (FinTech) companies. Loan disbursement is important for the FinTech industry in managing credit risk and increasing return on investment. K-Means Clustering and Complete Linkage are popular approaches in data clustering analysis. The research was conducted using loan datasets from FinTech customers processed by OJK. First, K-Means Clustering is used to group loan data and distribution of funds based on relevant attributes. Then, Complete Linkage is used as a comparison to evaluate the results of K-Means Clustering. Evaluation uses metrics such as Silhouette Score, Elbow method, and Davies-Bouldin Index. The results of the research are expected to provide insight for the FinTech industry in selecting more effective and efficient methods in classifying lending. DBI shows the lowest DBI (0.07364345) with K-Means Clustering, using 2 clusters. Research also shows that most regions in Indonesia comply with regulations, except DKI Jakarta.  
**Keywords** : K-means clustering, Complete Linkage, DBI, Silhouette score, Elbow method

## Abstrak

Penelitian ini bertujuan membandingkan efektivitas dan efisiensi metode *K-Means Clustering* dan *Complete Linkage* dalam mengelompokkan penyaluran pinjaman oleh perusahaan *Financial Technology (FinTech)*. Penyaluran pinjaman penting bagi industri *FinTech* dalam mengelola risiko kredit dan meningkatkan pengembalian investasi. *K-Means Clustering* dan *Complete Linkage* adalah pendekatan populer dalam analisis klusterisasi data. Penelitian dilakukan dengan dataset pinjaman dari pelanggan *FinTech* yang diolah oleh OJK. Pertama, *K-Means Clustering* digunakan untuk mengelompokkan data pinjaman dan penyebaran dana berdasarkan atribut relevan. Kemudian, *Complete Linkage* digunakan sebagai pembanding untuk mengevaluasi hasil *K-Means Clustering*. Evaluasi menggunakan metrik seperti *Silhouette Score*, *Elbow method*, dan *Davies-Bouldin Index*. Hasil penelitian diharapkan memberi wawasan bagi industri *FinTech* dalam memilih metode lebih efektif dan efisien dalam mengelompokkan penyaluran pinjaman. DBI menunjukkan DBI terendah (0,07364345) dengan *K-Means Clustering*, menggunakan 2 cluster. Penelitian juga menunjukkan sebagian besar daerah di Indonesia sesuai regulasi, kecuali DKI Jakarta.

**Kata kunci** : K-Means clustering, Complete Linkage, DBI, Silhouette score, Elbow method

## 1. Pendahuluan

Transformasi digital berdampak pada sektor keuangan **Error! Reference source not found.** Teknologi di bidang keuangan yang sering disebut sebagai *financial technology (fintech)* saat ini menarik perhatian para peneliti tidak hanya di bidang ekonomi dan bisnis, tetapi juga di bidang ilmu komputer, khususnya sistem informasi **Error! Reference source not found.** *Fintech* bersaing dengan layanan keuangan tradisional, menyediakan layanan yang berpusat pada pelanggan dan menggunakan teknologi internet untuk memfasilitasi akses **Error! Reference source not found.** *Fintech* saat ini telah memiliki banyak fungsi yang tidak hanya sebagai layanan transaksi keuangan online. Hasil riset Asosiasi *FinTech* Indonesia melaporkan bahwa saat ini perusahaan *fintech* di Indonesia masih didominasi oleh perusahaan pembayaran (44%), aggregator (15%), pembiayaan (15%), perencana keuangan untuk personal maupun perusahaan (10%), *crowdfunding* (8%), dan lainnya (8%). Terdapat enam model bisnis *fintech* yang diimplementasikan oleh pertumbuhan startup, yaitu pembayaran, *wealth management*, *crowdfunding*, *lending*, *capital market*, dan layanan asuransi **Error! Reference source not found.**

Hampir sepuluh kali lebih banyak platform *fintech* yang beroperasi secara legal di Indonesia mencoba beroperasi di luar regulasi karena banyaknya pemberian pinjaman yang salah sasaran dan banyak pemberi

pinjaman salah menyalurkan dana pinjaman. Oleh karena itu, OJK merupakan badan independen yang bertugas mengatur dan mengawasi jasa keuangan, termasuk praktik *Fintech Lending*. Regulasi OJK juga mengatur kewajiban pemberi pinjaman fintech dalam hal penilaian kelayakan kredit, pengelolaan risiko, pengendalian kualitas layanan, dan tindakan pemulihan kredit. *Fintech lending* diwajibkan untuk menyediakan informasi yang jelas dan akurat kepada konsumen, termasuk suku bunga, biaya, jangka waktu, dan risiko yang terkait dengan pinjaman, berdasarkan Peraturan Otoritas Jasa Keuangan (POJK) 77/POJK.01/2016 **Error! Reference source not found.** Berdasarkan peraturan tersebut, praktisi *fintech* yang terdaftar dan berlisensi harus memenuhi beberapa kriteria, antara lain: regulasi dalam pengawasan, bunga dan denda, kepatuhan terhadap peraturan, proses penagihan, bergabung dengan asosiasi, persyaratan pinjaman, layanan keluhan pelanggan yang tersedia, dan pembatasan akses ke data pribadi **Error! Reference source not found.**

Dari alasan tersebut maka akan dilakukan penelitian yang bertujuan untuk mengelompokkan penyaluran *financial technology* di Indonesia, pengelompokan menggunakan algoritma *clustering*. *Clustering* adalah proses membagi sekumpulan objek data menjadi sub-set yang disebut *cluster*. Objek di dalam sebuah *cluster* memiliki sifat yang mirip dan berbeda dengan *cluster* lainnya. Partisi dilakukan dengan menggunakan algoritma *clustering*, bukan secara manual. Oleh karena itu, pengelompokan sangat berguna untuk menemukan grup atau kelompok yang tidak dikenal dalam data. Algoritma ini memiliki kelebihan yaitu mudah diimplementasikan dan dijalankan, relatif cepat, mudah di kustomisasi, dan paling banyak digunakan dalam *data mining*. Pengelompokan ini adalah metode pengelompokan dokumen di mana mereka dikelompokkan berdasarkan kontennya untuk mengurangi ruang pencarian yang diperlukan untuk mengisi *query* **Error! Reference source not found.** Di Antara banyaknya *cluster*, terdapat dua metode pengelompokan dalam analisis *cluster* yaitu metode *hierarchical* dan metode *non-hierarchical*. Analisis klaster dengan metode *hierarchical* dilakukan dengan cara mengukur jarak kedekatan pada setiap objek yang kemudian membentuk sebuah *dendrogram*. Metode yang digunakan dalam *hierarchical* adalah metode *complete linkage*. Berbeda dengan metode *hierarchical*, metode *non-hierarchical* dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, proses pengelompokan dilakukan. Dalam metode *non-hierarchical*, cara analisisnya adalah *K-means Clustering*. Algoritma *K-Means* memiliki akurasi ukuran objek yang cukup tinggi, sehingga relatif lebih terukur dan efisien untuk memproses data dalam jumlah besar. Selain itu, algoritma *K-Means* tidak dipengaruhi oleh urutan objek **Error! Reference source not found.** Penelitian ini membandingkan metode untuk menentukan jumlah cluster yang tepat yaitu metode elbow method, koefisien silhouette, dan DBI. Dari latar belakang dan tinjauan serta melihat masalah diatas dan juga dari beberapa metode algoritma yang ada maka penelitian ini berjudul “Komparasi Antara Metode *K-Means Clustering* Dan *Complete Linkage* Dalam Pengelompokan Penyaluran Pinjaman Oleh *Financial Technology*”.

## 2. Metodologi

Penelitian ini menerapkan teknik *clustering*, *clustering* merupakan teknik *data mining* untuk mengorganisir data yang tidak memiliki label atau kelas. *Clustering* kadang-kadang disebut sebagai analisis *cluster*. *Clustering* dilakukan dengan membagi sejumlah besar data menjadi beberapa kelas atau *cluster* **Error! Reference source not found.** Pada dasarnya, *clustering* merupakan suatu metode dalam *data mining* untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data lainnya. Terdapat dua jenis *clustering*, yaitu *hierarchical clustering* dan *non-hierarchical clustering* [10].

Penelitian ini membandingkan teknik *Hierarchical clustering* dan *Non-hierarchical clustering*. *Hierarchical clustering* merupakan teknik pengelompokan yang membentuk hierarki atau struktur pohon berdasarkan level tertentu. Proses *clustering* dilakukan secara bertahap atau bertingkat. Biasanya, teknik ini digunakan untuk data yang tidak terlalu besar dan jumlah *cluster* yang akan dibentuk tidak diketahui sebelumnya [11]. Penelitian ini akan menggunakan metode *complete linkage* yang merupakan proses penggabungan dua objek atau lebih yang mempunyai jarak maksimum atau tetangga terjauh. Pada metode ini, jarak antara satu *cluster* dan *cluster* lain diukur berdasarkan jarak terjauh anggota-anggota mereka [12]. *Non-hierarchical cluster* menggabungkan atau memisahkan klaster untuk membentuk klaster baru, yang tidak mengikuti struktur seperti pohon pada pengelompokan hierarkis. Teknik pengelompokan data untuk memaksimalkan atau meminimalkan beberapa metrik dapat dianalisis dengan menggunakan teknik *clusterisasi non-hierarchical*, di mana jumlah *cluster* ditentukan secara manual [13]. Penelitian ini akan menggunakan algoritma *K-Means* yang merupakan metode analisis *cluster non-hierarchical* yang diawali dengan menentukan jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, selanjutnya

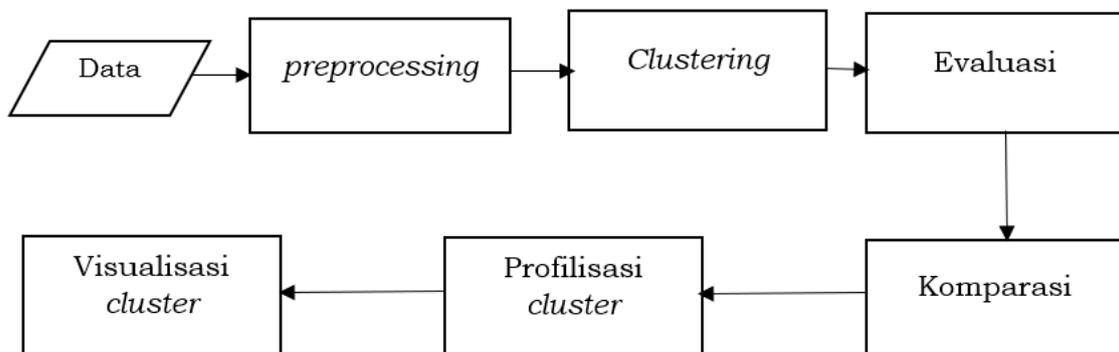
dilakukan proses *cluster* tanpa mengikuti proses *hierarchical* [14] . *K-Means* banyak digunakan di berbagai bidang karena sederhana dan mudah diimplementasikan. *K-Means clustering* sangat cocok untuk ukuran data yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan dengan metode *hierarchical*. Hasil pengelompokan dapat bergantung pada urutan pengamatan data [15] .

**2.1 Alur Penelitian**

Pada bagian ini akan dijelaskan mengenai alur penelitian pengelompokan data pinjaman dengan metode *K-means Clustering* dan *complete linkage*. Alur penelitian dijelaskan Langkah pertama menginputkan data yang diperoleh dari *statistic fintech* OJK. Langkah kedua adalah proses *pre-processing* dilakukan dengan normalisasi data menggunakan *z-scale*. Normalisasi data digunakan untuk menormalisasi nilai karena memiliki rentang nilai yang tinggi dan rendah sehingga bisa digunakan. Selanjutnya data akan masuk ke proses *clustering* dengan menggunakan metode yaitu *Hierarchical* dan *Non-Hierarchical clustering* dari hasil tersebut barulah dibandingkan. Langkah selanjutnya adalah proses pengujian metode dengan mempertimbangkan nilai *k* yang optimal. Setelah terpilih hasil *cluster* yang memiliki nilai validasi terbaik, maka hasilnya akan di visualisasi. Gambar 1 menjelaskan kerja sistem pada penelitian ini. Langkah pertama menginputkan data yang diperoleh dari *statistic fintech* OJK. Langkah kedua adalah proses *pre-processing* dilakukan dengan normalisasi data menggunakan *z-scale*. Normalisasi data digunakan untuk menormalisasi nilai karena memiliki rentang nilai yang tinggi dan rendah sehingga bisa digunakan. Selanjutnya data akan masuk ke proses *clustering* dengan menggunakan metode yaitu *Hierarchical* dan *Non-Hierarchical clustering* dari hasil tersebut barulah dibandingkan. Langkah selanjutnya adalah proses pengujian metode dengan mempertimbangkan nilai *k* yang optimal. Setelah terpilih hasil *cluster* yang memiliki nilai validasi terbaik, maka hasilnya akan di visualisasi dengan *scatterplot*.

**2.2 Data**

Data yang digunakan pada penelitian ini merupakan data yang dicatatkan oleh pihak Otoritas jasa keuangan langsung. Data yang digunakan adalah data statistik fintech lending periode bulan maret 2019 hingga Maret 2023. Data tersebut digunakan untuk membantu penelitian dalam mengklasterisasi daerah mana di indonesia yang paling banyak menggunakan jasa fintech lending untuk kebutuhan keuangan sehari-hari.



Gambar 1. Alur penelitian

Tabel 1. Data awal

Lokasi	Dec-22 Akumulasi Sejak Perusahaan Didirikan s.d Akhir Posisi Bulan	Jan-23 Akumulasi Sejak Perusahaan Didirikan s.d Akhir Posisi Bulan	Feb-23 Akumulasi Sejak Perusahaan Didirikan s.d Akhir Posisi Bulan	Mar-23 Akumulasi Sejak Perusahaan Didirikan s.d Akhir Posisi Bulan
a. Jawa	772.966	779.350	784.656	791.433

1. Banten	66.755	66.968	67.178	67.430
2. DKI Jakarta	298.244	302.654	306.371	310.820
3. Jawa Barat	202.226	203.086	203.744	204.666
4. Jawa Tengah	81.881	82.254	82.564	83.117
5. DI Yogyakarta	17.630	17.688	17.726	17.842
6. Jawa Timur	106.230	106.700	107.073	107.558

### 3. Hasil dan Pembahasan

#### Data Preprocessing

Data preprocessing adalah tahap kritis dalam proses analisis data, karena dapat mempengaruhi hasil akhir dari analisis atau model yang dibangun. proses yang fleksibel dan tergantung pada jenis data dan tujuan analisis atau model yang ingin dicapai. Preprocessing data terdiri dari beberapa tahap yaitu :

#### Data Cleaning

Data cleaning adalah memastikan bahwa data yang digunakan dalam analisis atau aplikasi lainnya adalah akurat, konsisten, lengkap, dan siap digunakan [16] . Kali ini data cleaning digunakan untuk mengecek missing values yaitu Nilai yang hilang atau kosong dalam data.

```

> anyNA(dataset)
[1] FALSE

> sapply(dataset, function(x) sum(is.na(x)))
      Lokasi / Locations
Jumlah Rekening Pemberi Pinjaman (akun) 0
Jumlah Dana yang Diberikan (miliar Rp)    0
      Jumlah Penerima Pinjaman (akun)    0
Jumlah Rekening Penerima Pinjaman Aktif (entitas) 0
      Outstanding Pinjaman (miliar Rp)    0

> summary(dataset)
Lokasi / Locations Jumlah Rekening Pemberi Pinjaman (akun)
Length:34          Min.      : 200
Class :character   1st Qu.:  872
Mode  :character   Median : 2206
                        Mean  : 238228
                        3rd Qu.:  5038
                        Max.   :7808096

Jumlah Dana yang Diberikan (miliar Rp) Jumlah Penerima Pinjaman (akun)
Min.      : 0.035                    Min.      : 12816
1st Qu.:  0.895                    1st Qu.:  39430
Median :  8.937                    Median : 109144
Mean   : 431.793                    Mean   : 421924
3rd Qu.: 18.768                    3rd Qu.: 242819
Max.   :12628.416                  Max.   :4013613

Jumlah Rekening Penerima Pinjaman Aktif (entitas)
Min.      : 22956
1st Qu.:  65858
Median : 167638
Mean   : 517631
3rd Qu.: 355699
Max.   :4813126

Outstanding Pinjaman (miliar Rp)
Min.      : 41.17
1st Qu.: 185.48
Median : 424.99
Mean   : 1500.58
3rd Qu.: 821.47
Max.   :13565.01
    
```

Gambar 2. Hasil Data Cleaning

Gambar 2 merupakan code serta hasil dari mengecek missing value. Dari studi kasus ini tidak terdapat missing value karena hasil menunjukkan false. Gambar 2 menampilkan sapply untuk mengecek adanya pengulangan data atau nilai object pada dataset untuk penelitian kali ini setelah dicek pada dataset ini tidak terdapat pengulangan data. Setelah melakukan pengecekan missing value dan sapply dan mendapatkan hasil tidak ada missing value dan tidak ada pengulangan kata maka pada Gambar 2 menampilkan summary atau ringkasan dari data yang digunakan pada penelitian. Pada summary menampilkan nilai minimum, median, mean, nilai maksimal, kuartier 1, dan kuartier 3 dari setiap attribute yang digunakan pada data penelitian ini.

#### Data Integration

Data integration adalah proses menggabungkan data dari berbagai sumber yang berbeda menjadi satu set data yang terpadu dan konsisten [17] . Tujuannya adalah untuk menciptakan representasi yang komprehensif dan holistik dari data yang dapat digunakan untuk analisis, pelaporan, dan pengambilan keputusan. Maka pada bagian ini dilakukan pengecekan variable, mengecek dimensi dataset.

```
-> #pengecekan variable
> dim(dataset)
[1] 34 6
> names(dataset)
[1] "Lokasi / Locations"
[2] "Jumlah Rekening Pemberi Pinjaman (akun)"
[3] "Jumlah Dana yang Diberikan (miliar Rp)"
[4] "Jumlah Penerima Pinjaman (akun)"
[5] "Jumlah Rekening Penerima Pinjaman Aktif (entitas)"
[6] "Outstanding Pinjaman (miliar Rp)"

> str(dataset)
tibble [34 x 6] (S3: tbl_df/tbl/data.frame)
 $ Lokasi / Locations           : chr [1:34] "Banten" "DKI Jakarta" "Jawa Barat" "Jawa Tengah" ...
 $ Jumlah Rekening Pemberi Pinjaman (akun) : num [1:34] 79203 7808096 64494 21119 6329 ...
 $ Jumlah Dana yang Diberikan (miliar Rp)   : num [1:34] 399 12628.4 810.2 71.1 17.4 ...
 $ Jumlah Penerima Pinjaman (akun)        : num [1:34] 1295053 2830474 4013613 1243163 256676 ...
 $ Jumlah Rekening Penerima Pinjaman Aktif (entitas): num [1:34] 1468180 2348600 4813126 1956868 394648 ...
 $ Outstanding Pinjaman (miliar Rp)       : num [1:34] 4420 10793 13565 3981 761 ...
```

Gambar 3. Hasil Data Integration

Gambar 3 menampilkan dimensi yang berisikan jumlah data dan jumlah attribute atau variable serta juga menampilkan variable yang digunakan yaitu: lokasi/locations, jumla rekening pemberi pinjaman(akun), jumlah dana yang diberikan (miliar Rp), jumlah penerima pinjaman (akun), jumlah rekening penerima pinjaman aktif (entitas), dan outstanding pinjaman (miliar Rp). Sedangkan Gambar 14 menampilkan struktur data.

**Transformasi data**

Transformasi data adalah proses mengubah atau memanipulasi data dari satu bentuk ke bentuk lain dengan tujuan tertentu. Tujuan dari transformasi data adalah untuk mengubah data menjadi lebih sesuai dengan kebutuhan analisis atau pemrosesan data yang akan dilakukan [18] . Pada sesi kali ini melakukan normalisasi data dan mengecek jarak matriks.

```
#normalisasi data
data<- dataset
unit_length <- function(x) {
  x / sqrt(sum((x^2)))
}
unit_length_df <- as.data.frame(lapply(data, unit_length))
head(unit_length_df)
View(unit_length_df)

> head(df)
# A tibble: 6 x 5
  Jumlah Rekening Pemberi Pinj_1 Jumlah Dana yang Dib_2 Jumlah Penerima Pinj_3
1 <dbl> <dbl> <dbl>
2 79203 399 1295053
3 7808096 12628.4 2830474
4 64494 810.2 4013613
5 21119 71.1 1243163
6 6329 17.4 256676
7 46058 487. 1636074

# $ abbreviated names: "Jumlah Rekening Pemberi Pinjaman (akun)",
# "Jumlah Dana yang Diberikan (miliar Rp)",
# "Jumlah Penerima Pinjaman (akun)",
# $ 2 more variables:
# "Jumlah Rekening Penerima Pinjaman Aktif (entitas)" <dbl>,
# "Outstanding Pinjaman (miliar Rp)" <dbl>

> ss <- sample(1:34, 34)
> df <- dataset[ss, ]
> df <- dataset[-c(3)]
> df.scaled <- scale(df)
> print(df.scaled)
  Jumlah Rekening pemberi Pinjaman (akun) Jumlah Dana yang Diberikan (miliar Rp) Jumlah Penerima Pinjaman (akun)
[1,] -0.1188810 -0.01515258 1.00983142
[2,] 5.6589282 5.64217667 2.78564708
[3,] -0.1298769 0.17504844 4.15402538
[4,] -0.1623023 -0.16686150 0.94981721
[5,] -0.1733587 -0.19167774 -0.19112008
[6,] -0.1436589 0.02563545 1.40424470
[7,] -0.1772595 -0.19923049 -0.43137931
[8,] -0.1655811 -0.18721999 0.04008093
[9,] -0.1758525 -0.19739954 -0.31517218
[10,] -0.1743223 -0.19086154 -0.27797355
[11,] -0.1752717 -0.19292862 -0.31120285
[12,] -0.1774486 -0.18891833 -0.44059482
[13,] -0.1764506 -0.19499438 -0.38980770
[14,] -0.1746699 -0.19549922 -0.15644738
[15,] -0.1775996 -0.19958043 -0.44297388
[16,] -0.1743275 -0.19468227 -0.20542911
[17,] -0.1747417 -0.19458773 -0.36707308
[18,] -0.1772901 -0.19733626 -0.41449113
[19,] -0.1778411 -0.19950972 -0.47223884
[20,] -0.1754743 -0.19405626 -0.24816191
[21,] -0.1768236 -0.19650389 -0.33138377
[22,] -0.1764312 -0.19833390 -0.35642921
[23,] -0.1778874 -0.19973161 -0.45615873
[24,] -0.1773858 -0.19950340 -0.43412152
[25,] -0.1779405 -0.19966318 -0.47316025
```

Gambar 4. Transformasi Data

Gambar 4 menampilkan code jarak antar matriks menggunakan metode *Euclidean*. *Euclidean distance* adalah jarak geometris yang paling umum digunakan. Jarak ini dihitung sebagai jarak linier (jarak lurus) antara dua titik dalam ruang multidimensi. Untuk dua matriks, jarak *Euclidean* dihitung sebagai akar kuadrat dari jumlah kuadrat perbedaan antara setiap elemen matriks. Metrik ini dapat digunakan untuk mengukur jarak antara titik-titik dalam ruang dimensi tinggi. Gambar 4 Menampilkan Hasil jarak matriks yang tepat. Itulah *Preprocessing data* untuk menormalisasi data sehingga data siap digunakan untuk proses *clustering* pada penelitian ini.

**Data Reduction**

*Data reduction*, atau pengurangan data, adalah proses mengurangi volume data yang tidak relevan, redundan, atau tidak perlu dalam suatu dataset. Tujuan dari pengurangan data adalah untuk menghilangkan kelebihan data yang dapat menyebabkan kompleksitas analisis yang tidak perlu,

mengurangi waktu dan sumber daya yang diperlukan untuk pemrosesan data, serta meningkatkan efisiensi dan kualitas analisis data. Pada data reduction kali ini akan melakukan KMO (*Kaiser-Meyer-Olkin*) dan PCA (*Principal Component Analysis*). KMO (*Kaiser-Meyer-Olkin*) adalah singkatan dari *Kaiser-Meyer-Olkin Measure of Sampling Adequacy*. KMO adalah metrik yang digunakan dalam analisis faktor atau analisis komponen utama (PCA) untuk mengukur kelayakan sampel data yang digunakan. Ini memberikan informasi tentang sejauh mana data yang digunakan cocok untuk analisis faktor atau PCA. kombinasi linear dari *attribute* awal menjadi beberapa komponen utama. Komponen-komponen utama ini dipilih berdasarkan kontribusinya terhadap variasi dalam data. Setelah proses KMO dan PCA, *attribute* yang tersisa adalah: Jumlah Rekening Pemberi Pinjaman (akun) Jumlah Dana yang Diberikan (miliar Rp) Jumlah Penerima Pinjaman (akun) Jumlah Rekening Penerima Pinjaman Aktif (entitas) Outstanding Pinjaman (miliar Rp). *Attribute* ini adalah representasi yang lebih ringkas dari data asli. Menggabungkan informasi dari berbagai *attribute* awal yang memberikan kontribusi terbesar terhadap variasi dalam data. Dengan menggunakan *attribute-attribute* yang sudah direduksi ini, Anda dapat menganalisis data dengan lebih efisien dan tetap mempertahankan informasi penting yang diwakili oleh komponen utama tersebut.

```
> kmo(rini[,1:5]) #sampel cukup untuk digunakan jika kmo>0.5
$KMO
[1] 0.7152648

$MSA
      Jumlah.Rekening.Pemberi.Pinjaman..akun.
      Jumlah.Dana.yang.Diberikan..miliar.Rp.
      Jumlah.Penerima.Pinjaman..akun.
      Jumlah.Rekening.Penerima.Pinjaman.Aktif..entitas.
      Outstanding.Pinjaman..miliar.Rp.
      PC1      PC2
Jumlah.Rekening.Pemberi.Pinjaman..akun.    0.43  0.61
Jumlah.Dana.yang.Diberikan..miliar.Rp.      0.44  0.49
Jumlah.Penerima.Pinjaman..akun.           0.45 -0.33
Jumlah.Rekening.Penerima.Pinjaman.Aktif..entitas. 0.45 -0.43
Outstanding.Pinjaman..miliar.Rp.          0.46 -0.30
```

Gambar 5. Hasil KMO dan PCA

Gambar 5 menampilkan yang termasuk ke dalam salah satu faktor dari 2 faktor yaitu PC1 dan PC2 bila nilainya berada di bawah -0.5 dan diatas 0.5. Bila ada yang tidak berada dalam batas maka dilihat dari nilai korelasi terbesar. Hasilnya dapat dilihat pada tabel dibawah ini dimana setiap variabel akan dominan pada salah satu komponen. Jadi komponen attribute yang digunakan adalah : Jumlah rekening pemberi pinjaman, Jumlah dana yang diberikan, Jumlah penerima pinjaman, Jumlah rekening penerima pinjaman aktif, dan Outstanding pinjaman. Setelah dilakukan preprocessing data maka didapatkan Tabel 2 sebagai data yang akan digunakan pada *clustering* untuk penelitian kali ini.

Tabel 2. Data bersih untuk penelitian ini

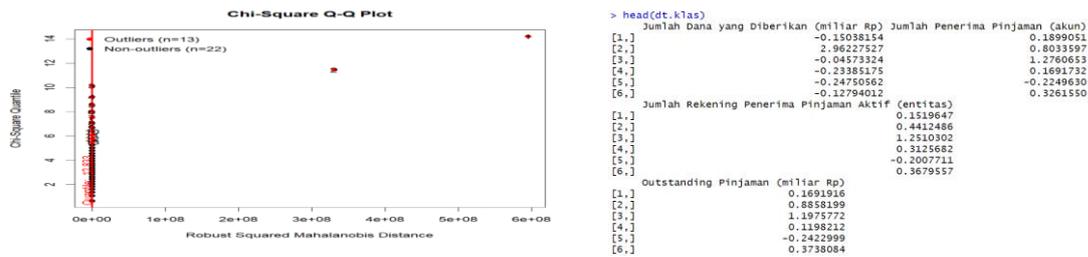
Lokasi	Jumlah Rekening Pemberi Pinjaman (akun)	Jumlah Dana yang Diberikan (miliar Rp)	Jumlah Penerima Pinjaman (akun)	Jumlah Rekening Penerima Pinjaman Aktif (entitas)	Outstanding Pinjaman (miliar Rp)
Banten	79.203	399,04	1.295.053	1.468.180	4.419,97
DKI Jakarta	7.808.096	12.628,42	2.830.474	2.348.600	10.792,67
Jawa Barat	64.494	810,19	4.013.613	4.813.126	13.565,01
Jawa Tengah	21.119	71,09	1.243.163	1.956.968	3.980,94
DI Yogyakarta	6.329	17,45	256.676	394.648	760,74
Jawa Timur	46.058	487,21	1.636.074	2.125.537	6.239,55

**Clustering**

Pada penelitian ini digunakan untuk mengklaster atau mengelompokkan pemberian dana pinjaman supaya penyebaran dana sesuai regulasi yang ada. Pada penelitian ini menggunakan 2 metode clustering yaitu K-Means Clustering dan Complete Linkage.

**K-Means Clustering**

Penerapan metode *K-means clustering* untuk menganalisis pola dan segmen penyaluran dana pinjaman. Tujuan dari analisis *K-Means clustering* ini adalah untuk memahami karakteristik kelompok pelanggan yang berbeda dan mengidentifikasi strategi pemasaran yang tepat untuk masing-masing kelompok. Hasil analisis K-means adalah pengelompokan data ke dalam K-Means cluster.



Gambar 6. Hasil Clustering K-Means Clustering

Gambar 6 menampilkan squer plot K-Means dengan 35 observasi yang mewakili setiap daerah yang ada di Indonesia, dan ada 13 nilai yang sangat tinggi atau sangat rendah dibandingkan dengan nilai lainnya, maka nilai-nilai tersebut dapat dianggap sebagai outlier. Sebaliknya, 22 nilai lainnya yang secara konsisten mengikuti pola atau tren yang diharapkan dianggap sebagai non-outlier. Setiap baris dalam tabel memberikan informasi serupa untuk observasi yang berbeda. Tabel ini dapat digunakan untuk melakukan analisis statistik, visualisasi, atau menerapkan metode clustering atau analisis lainnya tergantung pada tujuan Anda dalam pengolahan data tersebut.

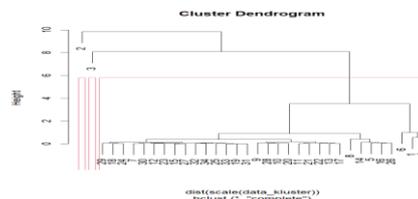
**Complete Linkage**

Metode pengelompokan data dalam hierarki klaster. Metode ini mengukur jarak antara dua kelompok dengan mencari jarak maksimum antara dua anggota dari kelompok yang berbeda. menampilkan nilai hasil analisis jika menggunakan complete linkage serta memberitahu metode apa yang dipakai, metode apa yang digunakan untuk menghitung jarak matrik, dan jumlah object data. Serta Pada dendrogram, sumbu vertikal menunjukkan tingkat kesamaan (jarak) antara objek atau klaster, sedangkan sumbu horizontal merepresentasikan objek atau klaster itu sendiri. Pada plot cluster dendrogram ini, garis vertikal yang menghubungkan cluster-cluster menunjukkan jarak di mana klaster-klaster tersebut digabungkan. Semakin panjang garis vertikal, semakin besar jarak atau perbedaan antara cluster-cluster yang digabungkan.

```
> corave
[1] 0.9930157
> hierarkiaive

Call:
hclust(d = dist(scale(data_kluster)))

Cluster method : complete
Distance       : euclidean
Number of objects: 34
```



Gambar 7. Hasil Clustering Complete Linkage

**Evaluasi hasil clustering**

Pada penelitian ini menggunakan 2 metode yaitu *K-Means clustering* dan *Complete Linkage* maka setelah melakukan clustering maka akan dilakukan evaluasi menggunakan nilai *Davies-Bouldin Index (DBI)* dan nilai *Elbow Method*.

**Davies Bouldin Index**

Salah satu metrik yang digunakan untuk mengukur kualitas partisi dalam analisis klaster (clustering). Metrik ini menggabungkan informasi tentang jarak antar titik dalam setiap klaster dengan jarak antara

klaster-klasternya. Indeks Davies-Bouldin dihitung berdasarkan prinsip bahwa partisi yang baik memiliki klaster yang kompak (titik-titik dalam klaster saling dekat) dan terpisah (klaster-klasternya berjauhan). menampilkan code untuk mengetahui pengelompokkan dengan menggunakan metode *K-Means Clustering*. Gambar 8 menampilkan hasil untuk perhitungan DBI jika dilihat dari jumlah pengelompokkan jika 2 *clustering* mendapatkan hasil 0,07364345, 3 *clustering* mendapatkan hasil 0,523967, 4 *clustering* mendapatkan hasil 0,3935374, dan 5 *clustering* mendapatkan hasil 0,4548166. Pada gambar 37 didapatkan hasil yaitu *cluster* optimal adalah dengan nilai DBI terendah yaitu 2 *cluster* dengan nilai 0,07364345.

```
> ##Melihat Nilai DB
> dbi_kmean2$DB
[1] 0.07364345
> dbi_kmean3$DB
[1] 0.523967
> dbi_kmean4$DB
[1] 0.3935374
> dbi_kmean5$DB
[1] 0.4548166

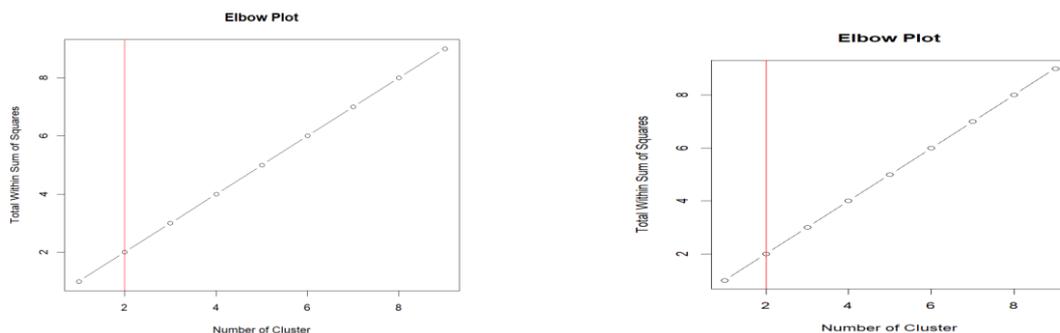
> # Menghitung Davies-Bouldin Index
> dbi <- davies_bouldin_index(data, cutree(hc, k = 2))
> # Menampilkan hasil
> print(dbi)
[1] 0.6567611
> # Menghitung Davies-Bouldin Index
> dbi <- davies_bouldin_index(data, cutree(hc, k = 3))
> # Menampilkan hasil
> print(dbi)
[1] 0.9533907
> # Menghitung Davies-Bouldin Index
> dbi <- davies_bouldin_index(data, cutree(hc, k = 4))
> # Menampilkan hasil
> print(dbi)
[1] 1.368071
> # Menghitung Davies-Bouldin Index
> dbi <- davies_bouldin_index(data, cutree(hc, k = 5))
> # Menampilkan hasil
> print(dbi)
[1] 1.807736
```

Gambar 8. Menampilkan hasil indeks *DBI* menggunakan *K-Means clustering* dan *Complete Linkage*

Gambar 8 juga menampilkan hasil DBI menggunakan metode *complete linkage* yang mendapatkan hasil jika 2 *cluster* nilai DBI nya adalah 0,6567611, 3 *cluster* nilai DBI nya adalah 0,9533907, 4 *cluster* nilai DBI nya adalah 1,368071, dan 5 *cluster* nilai DBI nya adalah 1,807736. Nilai DBI optimal adalah 2 *cluster* yaitu dengan nilai 0,6567611.

**Elbow Method**

*Elbow method* (metode siku) adalah teknik yang digunakan dalam analisis klaster (*clustering*) untuk membantu menentukan jumlah optimal klaster dalam sebuah dataset. Tujuan dari metode ini adalah untuk mencari "siku" pada grafik nilai inerti atau *sum of squared distances* yang dihasilkan oleh algoritma klastering terhadap jumlah klaster yang berbeda. menampilkan hasil *elbow method* yang menampilkan bahwa jumlah *cluster* optimal adalah 2 *cluster*.



Gambar 9. Hasil *elbow method* menggunakan *K-Means clustering* dan *Complete Linkage*

**Shillouette Coefficient**

Metode ini merupakan metode evaluasi *cluster* yang menggabungkan metode *cohesion* dan *separation*. *Cohession* diukur dengan menghitung seluruh objek yang terdapat dalam sebuah *cluster* dan *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah *cluster* dengan *cluster* terdekatnya. Pada penelitian kali ini menggunakan metode *K-Means Clustering* dan *Complete Linkage*. Gambar 10 menampilkan hasil *shillouette coefficient* menggunakan *K-Means clustering* beberapa *clustering* dan didapatkan hasil yang paling optimal adalah 3 *cluster* dengan nilai 0,5044519. Serta menampilkan hasil *shillouette coefficient* menggunakan *complete linkage* beberapa *clustering* dan didapatkan hasil yang paling optimal adalah 5 *cluster* dengan nilai 0,346174.

```
> # Menampilkan hasil koefisien siluet
> result <- data.frame(K = k_values, Silhouette = silhouette_values)
> print(result)
  K Silhouette
1 2  0.5533774
2 3  0.5044519
3 4  0.5159065
4 5  0.5088712
5 6  0.5965152

> groups <- cutree(hc, k = 5) # Ubah jumlah kluster sesuai kebutuhan
> # Menghitung koefisien siluet
> silhouette <- silhouette(groups, dist(data))
Warning message:
In dist(data) : NAs introduced by coercion
> # Menampilkan nilai rata-rata koefisien siluet
> mean_silhouette <- mean(silhouette[, "sil_width"])
> print(mean_silhouette)
[1] 0.346174
```

Gambar 10. Hasil *Sillhouette Coefficient* menggunakan *K-Means clustering* dan *Complete Linkage*

**Komparasi**

Komparasi antara hasil *clustering* menggunakan *complete linkage* dan *k-means clustering*. *k-means clustering* sebagai alternatif untuk menganalisis data *financial technology*. *K-means clustering* adalah metode *non-hierarchical* yang membagi data ke dalam kelompok-kelompok yang homogen berdasarkan pusat kelompok. Penerapan metode *complete linkage clustering* untuk menganalisis pola dan segmen peminjaman *financial technology*. Tujuan dari analisis ini adalah untuk memahami karakteristik kelompok penyaluran penyebaran dana yang berbeda menggunakan pendekatan *hierarchical clustering*. Dari hasil Evaluasi yang didapatkan menggunakan *Davies Bouldin Index*, *Elbow Method*, dan *Shillouette Coefficient* maka didapatkan hasil paling optimal yaitu:

Tabel 3. Hasil evaluasi semua indeks

Metode	DBI	Elbow Method	Shillouette Coefficient
K-Means Clustering	2 Cluster = 0,07364345	2 Cluster	2 Cluster = 0,5533774
	3 Cluster = 0,523967		3 Cluster = 0,5044519
	4 Cluster = 0,3935374		4 Cluster = 0,5159065
	5 Cluster = 0,4548166		5 Cluster = 0,5088712
Complete Linkage	2 Cluster = 0,6567611	2 Cluster	2 Cluster = 0,515983
	3 Cluster = 0,9533907		3 Cluster = 0,592692
	4 Cluster = 1,368071		4 Cluster = 0,4998128
	5 Cluster = 1,807736		5 Cluster = 0,346174

Dapat diketahui yang lebih optimal adalah 2 cluster dengan menggunakan metode *K-Means Clustering* mendapatkan nilai indeks DBI 0,07364345. Nilai Davies Bouldin Index (DBI) yang semakin mendekati nilai 0 menandakan semakin baik cluster yang diperoleh. Semakin rendah nilai DBI menunjukkan hasil cluster yang optimal[19].

**Profilisasi Clustering**

Dari evaluasi hasil maka akan digunakan metode *K-Means Clustering* dengan menggunakan 2 cluster menjelaskan berapa jumlah cluster yang akan digunakan dan apa keterangan dari setiap cluster serta terdapat perbedaan warna untuk setiap clustering. Jika sesuai regulasi maka akan berwarna biru sedangkan jika tidak termasuk regulasi maka akan berwarna hitam.

Tabel 4. *Attribute* dan nilai awal cluster

<b>Attribute</b>	<b>Cluster 1</b>	<b>Cluster 2</b>
Jumlah rekening pemberi pinjaman (akun)	8.839	7.808.096
Jumlah dana yang diberikan (miliar Rp)	62,19797	12.628,41568
Jumlah penerima pinjaman (akun)	348.938	2.830.474
Jumlah rekening penerimaan pinjaman aktif (entitas)	462.148	2.348.600,0
<i>Outstanding</i> pinjaman (miliar Rp)	1.219,004	10.792,673

Nilai centroid akhir juga dapat digunakan dalam visualisasi clustering. Pada nilai ini pada cluster 1 dilihat dari nilai setiap attribute yang dibandingkan antar attribute 1 dan lainnya memiliki nilai balance diatas 75 % maka dari itu cluster 1 adalah cluster yang sesuai regulasi berwarna biru sedangkan cluster 2 adalah cluster tidak sesuai regulasi berwarna hitam karena nilai balance dibawah 75% . Nilai centroid akhir dapat ditampilkan sebagai titik dengan warna atau simbol yang mewakili klaster tertentu. Ini membantu dalam memvisualisasikan lokasi dan perbedaan antara klaster-klaster tersebut. Otoritas Jasa Keuangan (OJK) menargetkan paling tidak 75 persen keseimbangan antar attribute yang tercatat pada layanan institusi *financial technology* yang ada supaya setiap provinsi yang ada di Indonesia termasuk dalam regulasi peredaran *financial technology* yang baik dan merata.

Tabel 5. Hasil persentase tiap daerah

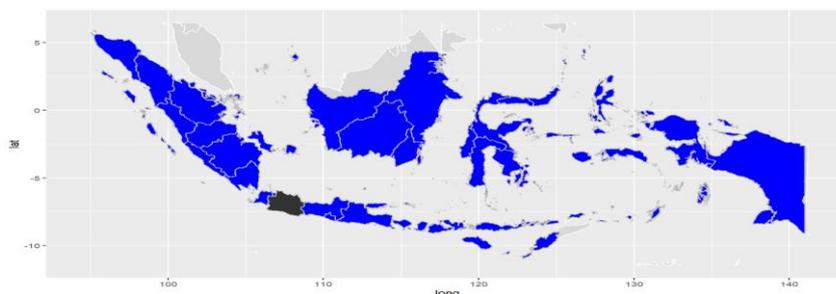
<b>Provinsi</b>	<b>Persentase (%)</b>	<b>Cluster</b>
Banten	87,44687	1
DKI Jakarta	23,22112	2
Jawa Barat	87,47456	1
Jawa Tengah	90,21916	1
DI Yogyakarta	90,58015	1
Jawa Timur	88,49047	1
Nangroe Aceh Darussalam	89,45461	1
Sumatera Utara	89,61696	1
Sumatera Barat	90,62297	1
Riau	90,14924	1
Kepulauan Riau	88,67365	1
Kepualauan Bangka Belitung	88,38539	1
Jambi	90,76388	1
Sumatera Selatan	89,32777	1
Bengkulu	90,31696	1
Lampung	89,29422	1
Kalimantan Barat	90,01896	1
Kalimantan Tengah	89,9877	1
Kalimantan Utara	90,94046	1

Kalimantan Timur	88,53401	1
Kalimantan Selatan	88,46152	1
Sulawesi Utara	90,36218	1
Gorontalo	87,23716	1
Sulawesi Tengah	91,37164	1
Sulawesi Barat	90,43865	1
Sulawesi Selatan	89,78508	1
Sulawesi Tenggara	91,28176	1
Bali	87,89225	1
Nusa Tenggara Barat	88,59927	1
Nusa Tenggara Timur	91,1541	1
Maluku Utara	91,12933	1
Maluku	90,61641	1
Papua Barat	90,79513	1
Papua	90,97231	1

Tabel 5 adalah hasil persentase dari setiap provinsi di Indonesia dapat dilihat bahwa setiap provinsi di Indonesia memiliki hasil yang sangat baik karena penyaluran dana pinjaman di Indonesia hampir di seluruh daerah sudah sesuai dengan regulasi yang telah dibuat oleh OJK. Daerah di Indonesia yang belum sesuai dengan regulasi OJK adalah DKI Jakarta dengan persentase 23,22112%.

#### Visualisasi Clustering

Visualisasi hasil clustering menggunakan metode *k-means*. Visualisasi ini bertujuan untuk memberikan pemahaman yang lebih intuitif mengenai pola dan segmen pelanggan dalam data *financial technology*. Pada visualisasi ini menggunakan 2 warna yaitu biru yang bermakna sesuai regulasi yang ada sedangkan warna hitam bermakna tidak sesuai regulasi.



Gambar 11. Hasil Visualisasi

Gambar 11 menampilkan hasil dari *clustering* menggunakan data *financial technology* di Indonesia yang dikeluarkan oleh otoritas jasa keuangan. Terlihat dari visualisasi hanya terdapat 1 daerah yang penyebaran *financial technology* yang tidak sesuai dengan kebijakan otoritas jasa keuangan menurut (POJK) 77/POJK.01/2016 (*PERATURAN OTORITAS JASA KEUANGAN*, n.d.) selain dari 1 daerah tersebut sudah sesuai dengan anjuran regulasi yang ada. Pengelompokan ini menggunakan 5 *attribute* dan dihitung sejak maret 2019 hingga maret 2023. Dari penelitian ini 1 daerah yang belum sesuai dengan regulasi yaitu DKI Jakarta selain dari provinsi itu sudah sesuai dengan regulasi penyebaran *financial technology* yang ada. Dilihat dari hasil, DKI Jakarta memiliki nilai-nilai *attribute* yang tidak sebanding diantaranya yaitu jumlah akun pemberi pinjaman adalah 7.808.096 sedangkan akun yang meminjam hanya 2.348.600 akun serta memiliki *outstanding* pinjaman yang lumayan besar yaitu 10.792 miliar.

#### 4. Kesimpulan

Berdasarkan hasil penelitian, ditemukan bahwa implementasi pengelompokan atau pengklasteran menggunakan algoritma *K-means* dan *Complete Linkage* pada *financial technology* (*fintech*) di Indonesia dapat membantu dalam mengidentifikasi kelompok-kelompok nasabah dengan kebutuhan dana pinjaman

serupa. Teknik klastering memungkinkan *fintech* untuk memahami dan menganalisis pola-pola dalam data nasabah guna meningkatkan keputusan terkait penyaluran dana pinjaman. Hasil dari pengklasteran menunjukkan bahwa menggunakan *complete linkage*, jumlah cluster optimal adalah 5 dengan nilai DBI sebesar 0,346174, sementara pada *k-means clustering*, jumlah cluster optimal adalah 2 dengan nilai DBI 0,07364345. Metode *K-Means Clustering* mampu menghasilkan kelompok yang jelas dan efisien dalam mengklasifikasikan data dengan variabel yang besar, sementara Metode *Complete Linkage* menghasilkan klaster yang kompak dan terpisah. Meskipun demikian, berdasarkan nilai DBI, 2 cluster menggunakan metode *K-Means Clustering* memiliki nilai terendah. Hasil penelitian juga menunjukkan bahwa hampir seluruh daerah di Indonesia telah mematuhi regulasi terkait penyaluran dana pinjaman, kecuali DKI Jakarta.

#### Daftar Pustaka

- [1] L. Zavolokina, M. Dolata, and G. Schwabe, "FinTech-What's in a Name?" 2016. [Online]. Available: <https://doi.org/10.5167/uzh-126806>.
- [2] K. Gai, M. Qiu, and X. Sun, "A survey on FinTech," *Journal of Network and Computer Applications*, vol. 103, pp. 262–273, 2018. [Online]. Available: <https://doi.org/10.1016/j.jnca.2017.10.011>.
- [3] P. Gomber, J. A. Koch, and M. Siering, "Digital Finance and FinTech: current research and future research directions," *Journal of Business Economics*, vol. 87, no. 5, pp. 537–580, 2017. [Online]. Available: <https://doi.org/10.1007/s11573-017-0852-x>.
- [4] I. Lee and Y. J. Shin, "Fintech: Ecosystem, business models, investment decisions, and challenges," *Business Horizons*, vol. 61, no. 1, pp. 35–46, 2018. [Online]. Available: <https://doi.org/10.1016/j.bushor.2017.09.003>.
- [5] PERATURAN OTORITAS JASA KEUANGAN. [Online]. Available: (n.d.).
- [6] FAQ: KATEGORI UMUM. [Online]. Available: (n.d.).
- [7] D. A. Grossman and O. Frieder, "Information retrieval," 2004.
- [8] Simamora, "Analisis multivariat," vol. 1. Gramedia Pustaka Utama, 2005.
- [9] J. Han, M. Kamber et al., "Designing Data-Intensive Web Applications," n.d.
- [10] B. Santosa, "Data mining: Teknik pemanfaatan data untuk keperluan bisnis," vol. 26. Graha Ilmu, 2007.
- [11] A. Tri, R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, 2019. [Online]. Available: <http://ejurnal.ung.ac.id/index.php/jjom.P->
- [12] Simamora and Bilson, "Analisis multivariat pemasaran," vol. 23. Gramedia Pustaka Utama, 25 C.E. [13] B. Wira, A. Budianto, and A. Wiguna, "IMPLEMENTASI METODE K-MEDOIDS CLUSTERING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI MAHASIWA BARU TAHUN 2018 DI UNIVERSITAS KANJURUHAN MALANG," *RAINSTEK: Jurnal Terapan Sains & Teknologi*, vol. 1, pp. 53-68, 2019. [Online]. Available: <https://doi.org/10.21067/jtst.v1i3.3046>.
- [14] Santosa, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis." Graha Ilmu, 2007.
- [15] Nugroho, "Pemrograman Java Menggunakan IDE Eclipse." Andi Publisher, 2008.
- [16] B. Cohen, D. K. Vawdrey, J. Liu, D. Caplan, E. Y. Furuya, F. W. Mis, and E. Larson, "Challenges associated with using large data sets for quality assessment and research in clinical settings," *Policy, Politics, & Nursing Practice*, vol. 16, no. 3-4, pp. 117-124, 2015.
- [17] D. Henderson, "Overcoming Data Integration Problems," *InfoManagement Direct*, 1998. [Online]. Available: <http://www.informationmanagement.com/infodirect/19980701/924-1.html>. Retrieved December 2, 2010.
- [18] R. R. Suryono, I. Budi, and B. Purwandari, "Challenges and trends of financial technology (Fintech): A systematic literature review," *Information (Switzerland)*, vol. 11, no. 12, pp. 1–20, 2020. [Online]. Available: <https://doi.org/10.3390/info11120590>.
- [19] M. F. Fahmi Wirawan and Y. K. Suprpto, "Segmentation and distribution of watershed using K-Modes clustering algorithm and daviesbouldin Index based on geographic information system (GIS)," in *Proceedings 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, 2017.